European projects update

Dan Ellis
International Computer Science Institute, Berkeley CA
<dpwe@icsi.berkeley.edu>

Outline

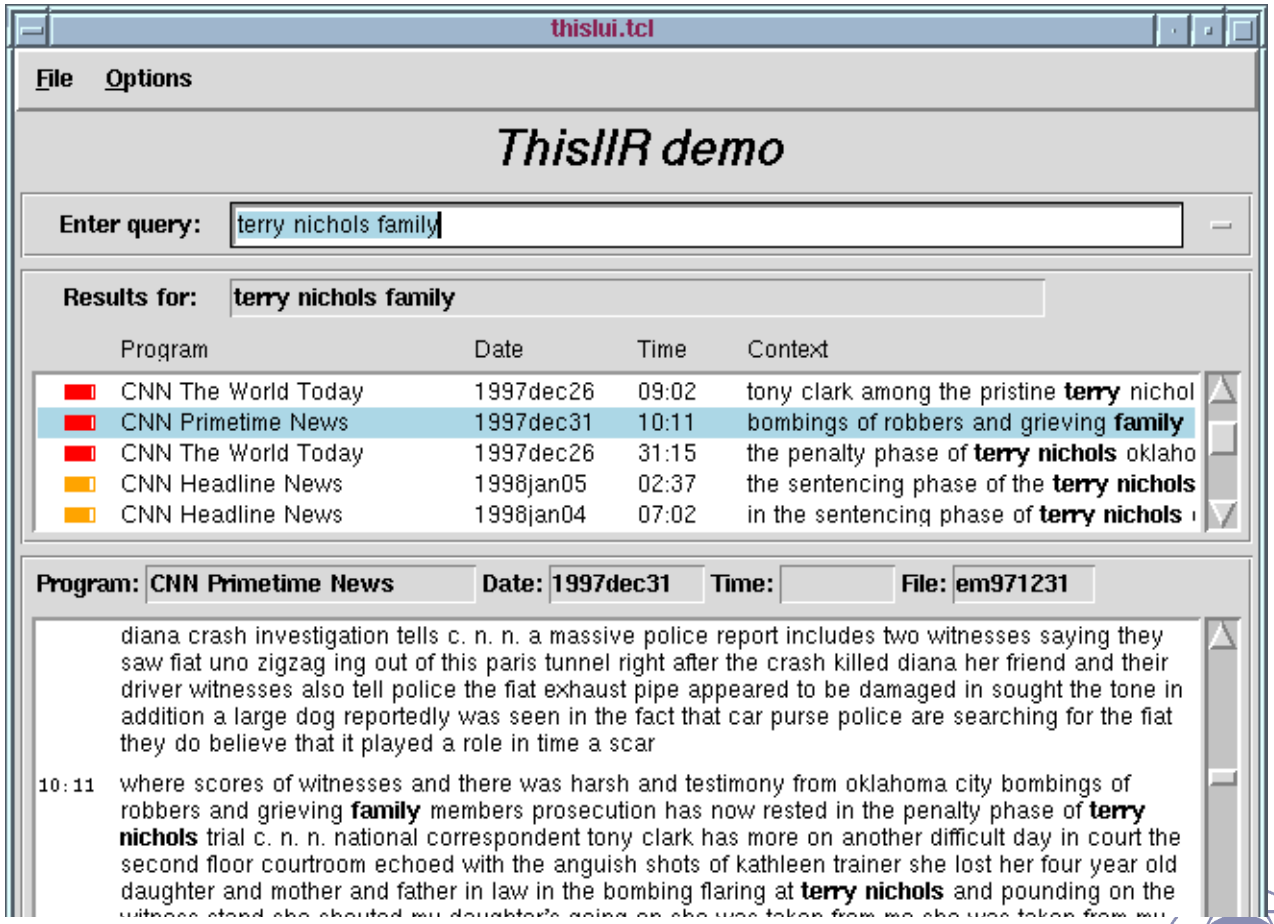
1. ***Thisl*** **final year planning meeting (BBC, London, feb03)**
2. **Future PDA project discussion (Sheffield Univ., feb04)**
3. ***RESPITE*** **kickoff meeting (ICP, Grenoble, feb07-08)**

Thisl final year planning meeting

- **Thisl project:**
  Using ASR (&c) to index BBC news archives
- **ESCA workshop on Spoken Document Retrieval (SDR) - April, Cambridge**
  - systems, IR/IE
  - demos, including thislIR
- **Current actions:**
  - finalize UKEng system to run on 1000hr archive (to demo in April)
  - ICSI to train MLP on BBC data (50hr)
  - segmentation? speaker tracking?
- **Other highlights:**
  - Latent Semantic Analysis with Self-Organizing Maps for SDR (Mikko Kurimo/IDIAP)
  - Confidence-based embedded training (Tony)

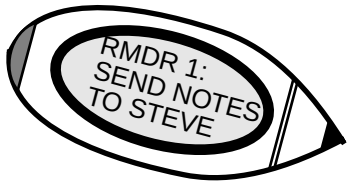
Thisl demo

- **Stand-alone Tcl/Tk implementation**
  - doesn't require httpd
  - speech-input ready



PDA proposal discussion

- **New EU 'Future & Emerging Tech' funding program (Framework 5)**
  - meeting to plan a possible project proposal
- **Partners interested in speech-centered PDA:**
  - Roger Tucker (HP Labs): audio info appliances
  - Tony Robinson (Cambridge): ASR, systems
  - Steve Renals (Sheffield): information access (IA)
  - IDIAP: multi-modal, Hervé's 'encapsulators'
  - FPMs: synthesis
  - ICSI: robust SR, UI/apps
- **HP vision**
  - 'fat pen' with mic & small display
  - dictaphone meets PDA
  - docks to PC
  - use pen-motion info?
  - access: skimming, summary, keyword search...



SpeechPDA proposal (cont'd)

- **Current application definition**
  - palm-style machine, docks, mixed online/offline
  - ASR for search, control, form-filling, summaries
  - voice notes/dictation as primary focus; 'ambient' recording as parallel track (shares IA)
- **Work packets**
  - ASR: wordspotting / robustness / efficiency
  - Info access: browse / skim / structure / SDR
  - System integration & user interface
  - ? other modalities (stylus, video?)
  - Evaluation: components / integrated system
- **Other observations**
  - meeting recorder doesn't have to be palm-size (hierarchy of size/power trade-offs?)
  - using ASR (IA) independent of where it's done
  - OK to have algorithms without a real prototype

RESPITE kickoff meeting

- **REcognition of Speech by Partial Information TEchniques**
  - Auditory scene analysis etc. to find information
  - Multistream & missing-data to exploit it
  - new 3yr EU-funded project:
    Sheffield, IDIAP, FPMs, ICP Grenoble, ICSI, DaimlerChrysler, Matra
- **Rationalize work at partner labs:**
  - missing-data at SU, IDIAP, FPMs
  - multi-stream at IDIAP, ICSI, ICP
- **Baseline task:**
  - “Aurora” Distributed Speech Recognition task: TIDIGITS corrupted in various ways
  - HTK and/or comparable system configuration
- **CASA toolkit**
  - practical information for use in ASR

Issues in missing data (Sheff/IDIAP)

- **Input features tagged as present/missing**
  - e.g. by subband SNR, scene analysis
- **Classic: 'Class imputation'**
  - integrate over missing data dimensions to evaluate output likelihoods:
    $p(X|q) = \int p(X_{present}, X_{missing} | q)\, dX_{missing}$
  - i.e. just skip dimensions of Gaussian
  - can use 'upper bounds' on spectral values
- **New: 'Data imputation'**
  - use $E[X_{missing} | X_{present}, q]$
  - permits cepstra, deltas
- **What about connectionist systems?**
  - also permitted by data imputation
  - or Radial Basis Function neural networks? (Andy Morris/IDIAP)

Harmonicity labelling for multistream

(Herve Glotin, ICP/IDIAP)

- **'Pitch pulse' in envelope autocorrelation is correlated to subband SNR (for vowels)**
- **Use artificial mixtures to train $R_{xx} \rightarrow$ SNR map**
- **'Full combination' multistream needs weights:**
  - $p(q \mid a,b,c,d) = \sum_S p(S) \cdot p(q \mid S,a,b,c,d)$

    S ranges over 16 possible combinations
  - uniform weighting is worse than best single S
  - $p(S) = p(\text{SNR} > \theta)$ gives best result: NB-noise-Num95: 15%FB $\rightarrow$ 13.3%WMB

Multistream vs. alternatives

(Andy Morris)

- **Keep an eye on alternative techniques**
  - e.g. noise robustness through spectral subtraction, microphone techniques
- **Techniques may not combine additively**
  - e.g. log Rasta vs. j-Rasta for full/multi-band: